## **Communications**

Part V of *Human Papillomaviruses 1995* is divided into two sections. The first introduces ways of accessing papillomavirus information from the World Wide Web (WWW), the second is a list of references.

### Papillomavirus Information on the World Wide Web

All of the information in *Human Papillomaviruses 1995* and its predecessor in 1994, as well as numerous data files, are available to the public through the World Wide Web, a system of links between Internet locations that can be navigated with graphical "browsing tools" such as Mosaic and Netscape. We are strenuously urging our subscribers to make use of our Web site. Using electronic documents and data offers many advantages over the production and use of paper documents.

- It saves paper, ink, metal, solvents, plastics, postage, fossil fuels, and time that go into the printing and distribution of hundreds of copies of this large book.
- Electronic documents that contain all the formatting and complex graphics of a printed document can now be produced and placed on the Web where users may view and print them. For example, an exact replica of this entire document, *Human Papillomaviruses 1995* is available from our Web site.
- Electronic documents are readily corrected and kept up to date. As we prepared sections of this year's compendium we have been "prepublishing" them on our Web site. In this way the latest and most accurate information is always available to users.
- There is also the possibility of interaction between users and the database staff at Los Alamos. Just as documents and data can be retrieved from the Web site, so can they be posted to the site.
- Data and documents on the Web, far from being static pieces of paper, are embedded in a system that allows searching for items of interest—references, sequences, author names, etc. Our search capabilities are relatively primitive now but will become more powerful in the future.
- The Web, as its name implies, is a system of links spread throughout the world. Our Web site provides links to other sites that are useful to virologists and sequence biologists.

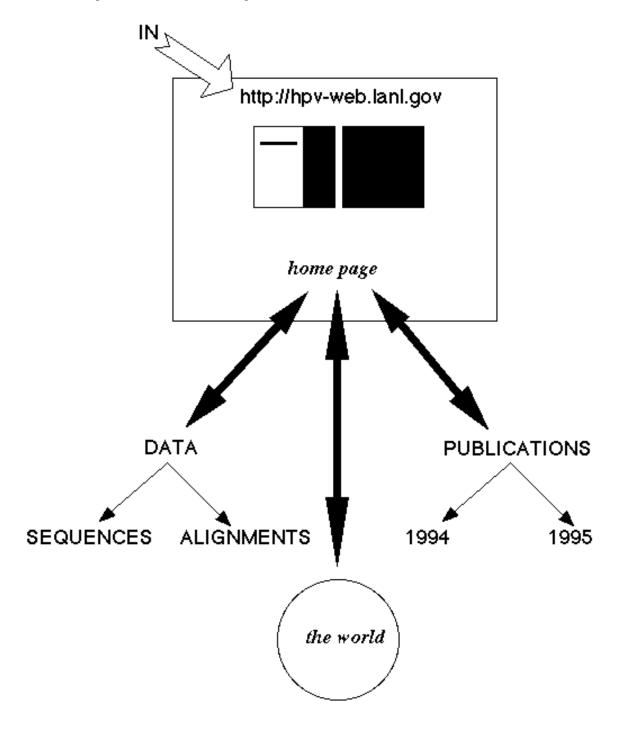
In order to access the WWW, you must have, (1) a windowing capability such as X Windows, MS Windows, or Apple Macintosh, (2) an Internet Protocol (IP) connection, and (3) a copy of a browser program such as Mosaic or Netscape for the appropriate windows environment. If you are not already using Mosaic (a free program) you can get a copy by anonymous ftp from ftp.ncsa.uiuc.edu. Netscape may be purchased by sending e-mail to sales@netscape.com or phoning 415-528-2555. Our use of Netscape in the following description of our Web site is purely illustrative and should not be construed as an endorsement of that product.

The browsing tools designed for the WWW make it easier than before to locate and acquire information on the Internet because, instead of typing arcane commands into your computer, you simply point and click at text and objects on your screen. These objects are either links to files, which cause the files to be sent to your local computer, or they are links to other machines from which files may be accessed. "Files" can mean words, pictures, sounds, or movies. The WWW address (or "URL") for the HPV database is:

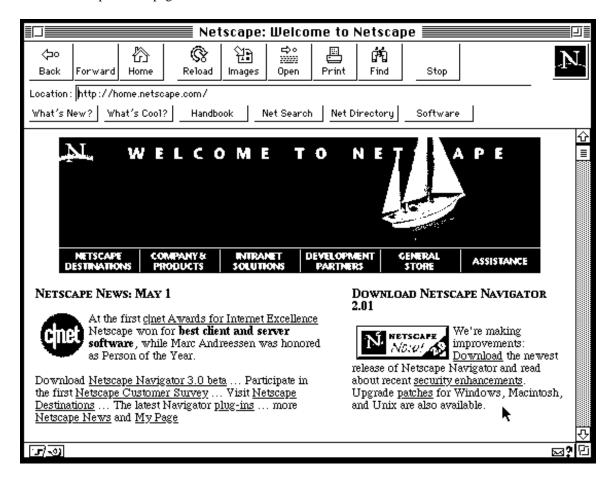
http://hpv-web.lanl.gov

**Tour of the World Wide Web** We present here a short tour of our Web site, which may give the uninitiated a feel for how the Web works.

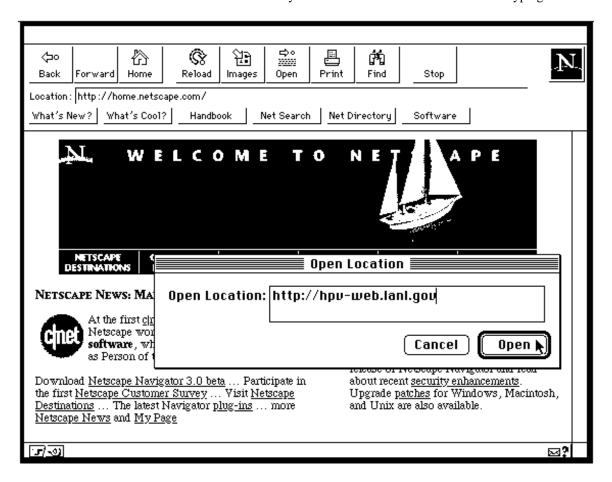
Our site on the Web is diagrammed as in the illustration below. When you access our site from elsewhere (represented by the red "IN" arrow) you are at our so-called "home page." From here you can follow links to our data, our publications, or other sites spread across the world.



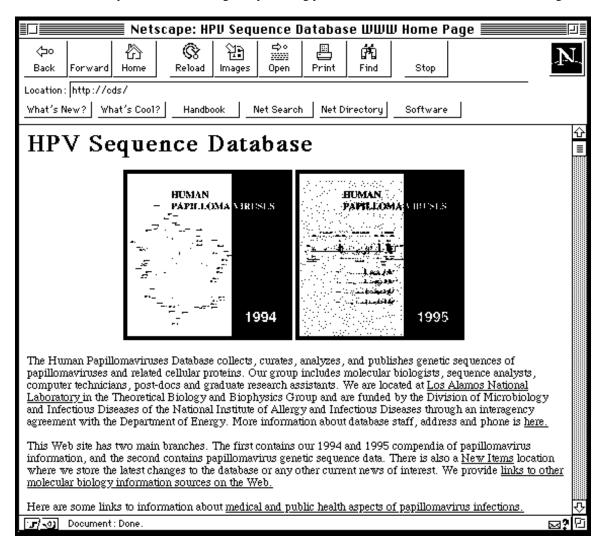
You begin by "launching" your browser program. (The browser in the illustrations here is Netscape.) In the Netscape window on your screen there appears the starting point for your navigation of the Web (Screen 1). This starting point can be any address you have specified, it could even be the HPV database URL; by default it is the Netscape "home page."



From your starting point you select "Open Location" from the File menu in the upper left corner of the window. A small window entitled Open Location appears, you type in the URL of the HPV Database http://hpv-web.lanl.gov in the space provided (Screen 2), and click Open. You can also store frequently visited URLs in a list called "bookmarks" or "hotlist" and select your destination from this list rather than typing its URL.



After a few seconds a new image appears in the window (Screen 3), the home page of the HPV Database. The text in the window is formatted and pleasantly readable. All the text on the home page cannot fit on the screen at once, but you can scroll through it by clicking your mouse on the down arrow in the lower right.



There are two branches to our site: (1) publications and (2) data. We will follow the data branch first. As you scroll down through our home page you come to an area labeled "II. Data" as shown here. The data area of this Web site contains links to genetic sequences of papillomaviruses stored in GenBank, EMBL, and SWISS-PROT formats. It also contains links to alignments of papillomavirus genes and their proteins. These data files are stored on our ftp server, atlas. You can follow any of these links to that server and look there for any data you require. Another approach to finding the data you need would be to use the HPV Sequence Search mechanism, a link about to be opened by the pointing hand in the graphic above.

### II. Data

Besides the compendium, we also provide **genetic sequence information** for papilloma viruses and related cellular proteins.

Attention Macintosh Netscape users.

### Sequences

- Read Me
- Database Map (Viewable Jpeg image.)
- 🗣 <u>Database Map (PDF image)</u>
- The HPV Sequence Search Interface
- GenBank № hmatted Nucleic Acid Sequences
- EMBL Formatted Nucleic Acid Sequences
- SWISS-PROT Formatted Probable Protein Sequences

Upon clicking HPV Sequence Search the search area appears on the screen. An explanation of how to use the search utility is included in the window. We type in a search pattern – in this example U12499, a GenBank accession number – and hit the carriage return.

This	iз	a searchable	index.	Enter	search	keywords:	U12499

## **HPV Database Search**

Enter your search in the search field above.

In this window you can search for sequences by typing key words into the space above. Searchable items include any words that might be included in the GenBank file categories of LOCUS, DEFINITION, ACCESSION number, KEYWORDS, SOURCE, REFERENCE, AUTHORS, TITLE, and JOURNAL. Note, however, that COMMENTS, FEATURES, and the nucleotide sequence itself are not currently available for search.

The search allows for partial matches with the search word. Thus searching for HPV3 will find HPV3, HPV30, HPV35h, etc. If you wish to search for multiple words separate them by a space. But be aware that the search is an 'or' search not an 'and' search. If you are interested in finding a sequence whose authors were Cole and Danos, and so type Cole Danos into the search box, you will receive not just one answer, but all the GenBank files that mention either Cole or Danos. Still, multiple query words would be useful for retrieving multiple files based on unique characteristics such as Accession numbers.

The result of the search, an abbreviated GenBank file, appears. We can retrieve the entire file by positioning the pointing hand on the link called Sequence and clicking.

#### Result of search for "U12499". LOCUS HPV62MY911 449 bp ds-DNA 16-0CT-1994 VRL DEFINITION Human papillomavirus type 62 (HPV-62), partial L1 cds, My09/My11 region. ACCESSION U12499 Human papillomavirus type 62 DNA isolated from a patient with SOURCE vulvar intraepithelial neoplasia (VaIN). ORGANISM Human papillomavirus type 62 Viridae; ds-DNA nonenveloped viruses; Papovaviridae; Papillomavirus. REFERENCE 1 (bases 1 to 449) Bernard, H.-U., Chan, S.-Y., Manos, M.M., Ong, C.-K., Villa, L.L., Delius, H., Peyton, C.L., Bauer, H.M., and Wheeler, C.M. AUTHORS TITLE Identification and assessment of known and novel human papillomaviruses by PCR amplification, restriction fragment length polymorphisms, nucleotide sequence, and phylogenetic algorithms J. Infect. Dis. (1994) In press JOURNAL <u>Sequence</u> 1 entries were found

The result, the complete GenBank file, is shown as a cropped image in the illustration below. This file may be transferred to your local computer in its entirety by clicking on "File" in the menu bar at the top of the screen and selecting "Save As . . . " from that menu. (Note! users of Netscape, for Macintosh, at least, may have to use a slightly different procedure for downloading files. There is a link entitled "Attention Macintosh Netscape users" on our home page. Consult that for detailed instructions.) Our search mechanism is currently quite primitive. You can only search for GenBank files. If you are interested in EMBL-format files you will need to find them "by hand" on atlas.

```
LOCUS
             HPV62MY911
                            449 bp ds-DNA
                                                         VRL
                                                                   16-0CT-1994
DEFINITION
            Human papillomavirus type 62 (HPV-62), partial L1 cds, My09/My11
             region.
ACCESSION
             012499
SOURCE
             Human papillomavirus type 62 DNA isolated from a patient with
             vulvar intraepithelial neoplasia (VaIN).
  ORGANISM Human papillomavirus type 62
             Viridae; ds-DNA nonenveloped viruses; Papovaviridae;
             Papillomavirus.
REFERENCE
             1 (bases 1 to 449)
  AUTHORS
             Bernard, H.-U., Chan, S.-Y., Manos, M.M., Ong, C.-K., Villa, L.L.,
             Delius, H., Peyton, C.L., Bauer, H.M., and Wheeler, C.M.
  TITLE
             Identification and assessment of known and novel human
             papillomaviruses by PCR amplification, restriction fragment
             length polymorphisms, nucleotide sequence, and phylogenetic
             algorithms
             J. Infect. Dis. (1994) In press
  JOURNAL.
  STANDARD
             full staff_review
             HPV-62 has recently been characterized and isolated from a vulvar
COMMENT
            intraepithelial neoplasia by Dr. T. htsukura. The cloned DNA was subsequently sequenced by Dr. H. Delius over the L1 MY09/MY11
             segment. HPV-62 and the several other types recently sequenced
            over this region by Dr. Delius were used as type-specific probes to screen DNA for novel genital HPV types. The screened DNA was
             obtained from four recent epidemiological studies. Primer regions
             are annotated in the sequence; information in this region is not
             accurate due to primer degeneracy.
FEATURES
                       Location/Qualifiers
                       <1..>449
                       /note="putative"
                       /product="major capsid protein"
                       /gene="L1"
                       /note="putative"
                       /codon_start=1
                       translation="AQGHNNGICWFNELFVTVVDTTRSTNFTICTASTAAAEYTATNF/
                       REFLRHTEEFDLQFIFQLCKIQLTPEIMAYLHNMNKDLLDDWNFGVLPPPSTSLDETY
                      HYFESRAITCQRGLPTRPKVDPYAQMTFWTVDLKDKLSTDLDQFPLG"
BASE COUNT
                 116 a
                                      99 q
                                               142 t
ORIGIN
        1 gcacagggtc ataataatgg tatttgttgg tttaatgaac tgtttgttac tgtggtggat
       61 actaccagaa gtactaattt tactatttgt accgcctcca ctgctgcagc agaatacacg
      121 gctaccaact ttagggaatt tttgcgacac acggaggaat ttgatttgca atttatattt
```

If we return to the data area of the home page (II. Data) we can follow links to alignment files.

## II. Data

Besides the compendium, we also provide genetic sequence information

Attention Macintosh Netscape users.

# Sequences

- Read Me
- Database Map (Viewable Jpeg image.)
- Database Map (PDF image)
- The HPV Sequence Search Interface
- GenBank Formatted Nucleic Acid Sequences
- EMBL Formatted Nucleic Acid Sequences
- SWISS-PROT Formatted Probable Protein Sequences

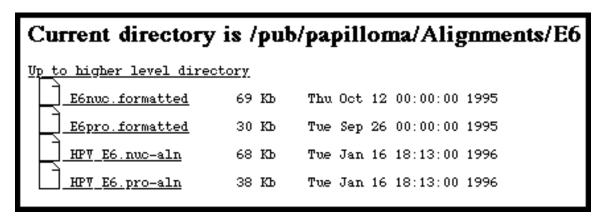
# Alignments

• <u>Nucleic Acid and Protein Alignments</u>

This is what the alignments directory on atlas looks like. There is a folder icon for each gene arrayed along the left edge of the page. There is also a fairly complete ReadMe file that discusses the alignment files contained within the folders. If we are interested in the E6 alignment we would open that folder as the pointing hand is about to do.

Current directory is /pub/papilloma/Alignments						
Up_to higher level directory						
Complete Genomes/		Fri Fe				
		Fri Ja				
		Mon Oc				
		Fri Ja				
		Mon Oc				
		Mon Oc				
		Mon Oc				
<u> L1/</u>		Mon Oc				
L1 consensus primer region/		Mon Oc				
<u></u>		Mon Oc				
LCR/		Mon Oc				
Miscellaneous/		Mon Ju				
ReadMe Alignments	2 Kb	Fri Ja				

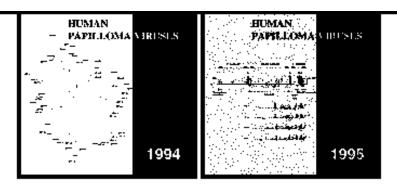
Once inside the E6 folder we see there are two pairs of files. The files with "formatted" in their names are the formatted nucleotide and protein alignments one sees in printed publications; the other two files are the alignments in "Intelligenetics" format.



The latter, unformatted files, are what you would want to download if you are interested in manipulating the alignments or adding your own sequences to an existing alignment. One such file, the HPV\_E6.pro-aln amino acid alignment is shown below. This ends our introduction to the *data* resources of the Web site.

```
Modified by adf on Tue Sep 19 12:14:13 1995
  Modified by cxc on Tue Sep 19 10:44:19 1995
;; Modified by adf on Fri Sep 15 13:57:59 1995
; Nucleic Acid Consensus Sequence
; for sequences 1-40
; Modified by adf on Fri Se 15 13:35:56 1995
 Written from MASE, from file "temp.21019"
SuperA.com
MKKGVTENGATESGAYKKLLKLW??m?m???f?????pr?l??Lo??????1?1?1?Cv?Ckk?1???
evy?fa??dl?vvyr?g-?pyaaC??Cl?f??k?r?yrhy?ys?yg?t?e??t???l??lliRC??C?kP
L?p?eK?rh???k?rfh?i?g?w----t6?C??Cw???????????et?v??DQQGSSFLQA*****
**********
 Modified by adf on Mon Aug 28 12:12:59 1995
HP754
*****************************MSATEPHTDQPRTLADLCKYCNIPMHSLQLPCAFCKKTYCTA
EIYAFOYKDLFVVWRHG-FPHAACALCLELHGOINYRRHRDRACLWETVEOECGKPLEEIFIRCWLCHKP
LCNVEKQRHVDYNRRFHCVRGYW----KGRCLHCWKP*
BLANK
; Nucleic Acid Consensus Sequence
 for sequences 1-2
 Modified by adf on Fri Sep 15 13:35:39 1995
 Written from MASE, from file "temp.21019"
GroupA1.com
E??AYHFKDL?VVW?K?-FPYAACAFCLEF?SK?CALRHY?RSAFW?TVE?ETGLLLEEQ?IRCA?CQKP
LS?SEK?HHI??G??F?FIL??W----TGRCT?CR?OCVERRLP***********************
```

Let us now return to the HPV home page and follow the links that lead down the other main branch of this site, the "on-line version of the *Human Papillomaviruses compendia*." There are various routes we can take to these publications. We can click the orange or green book covers at the top of the home page or we can follow the links under part "I. The HPV Compendium," also on the home page.



The Human Papillomaviruses Database collects, curates, analyzes, and publishes genetic sequences of papillomaviruses and related cellular proteins. Our group includes molecular biologists, sequence analysts, computer technicians, post-docs and graduate research assistants. We are located at <u>Los Alamos National Laboratory</u> in the Theoretical Biology and Biophysics Group and are funded by the Division of Microbiology and Infectious Diseases of the National Institute of Allergy and Infectious Diseases through an interagency agreement with the Department of Energy. More information about database staff, address and phone is <a href="here">here</a>.

This Web site has two main branches. The first contains our 1994 and 1995 compendia of papillomavirus information, and the second contains papillomavirus genetic sequence data. There is also a <u>New Items</u> location where we store the latest changes to the database or any other current news of interest. We provide links to other molecular biology information sources on the Web.

Here are some links to information about medical and public health aspects of papillomavirus infections.

We appreciate receiving your comments, queries, and criticisms about this Web site.

### I. The HPV Compendium

Each year we publish a compendium of papillomavirus information called <u>Human Papillomaviruses: A Compilation and Analysis of Nucleic Acid and Amino Acid Sequences.</u> which can shaw be downloaded from this Web site. The 1995 Compendium was released on October 26. Included in the Compendium are

Either way we arrive at the page displayed below. This page introduces the on-line version of our publication and describes the Adobe Acrobat viewer which you will need to view the publications. The viewer is free and can be obtained by following the link shown below.

## The Human Papillomaviruses Compendium On Line

You may now view and print out all or parts of the *Human Papillomaviruses 1994* Compendium or the *Human Papillomaviruses 1995* Compendium that interest you from this Web site. These two documents consist of

Adobe Acrobat PDF files. "PDF" (Portable Document Format) files retain all the typographical qualities of the original document -- fonts, tables, illustrations. They can be viewed and printed by anyone using a freely available copy of the **Adobe Acrobat** viewer which is available for Macintosh, UNIX, and DOS computers. Your free copy of the Acrobat viewer is available from Adobe Systems.

If you wish to learn more about our use of Acrobat and PDF files click here.

### Hard Copy

The Compendium has been published in paper form and is available to qualified researchers who cannot make use of the electronic hard copy provided here. If you really need a complete paper copy write to us at

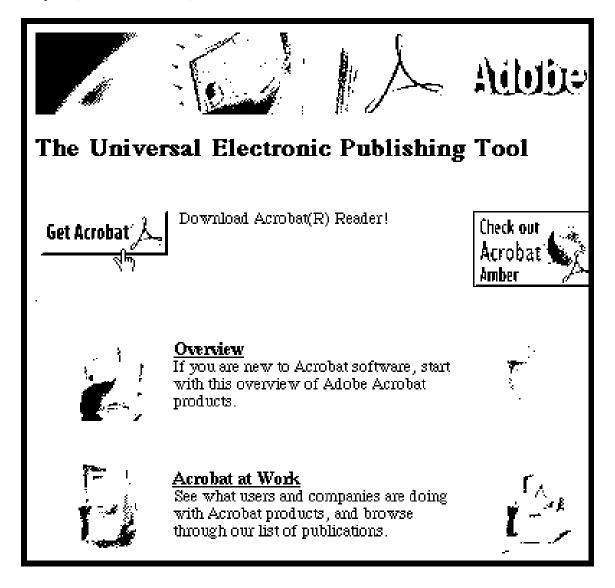
HPV Database Mail Stop K710 Los Alamos National Laboratory Los Alamos, NM 87545 USA

or by sending us e-mail at cxc@transposon.lanl.gov

1994

CONTENTS

This is what the Adobe Acrobat page looks like. You can download the Acrobat viewer appropriate to your computer (Mac, DOS, or UNIX) here.



Returning to the compendium introduction at our site (below) we can view the tables of contents of the 1994 or 1995 *Human Papillomaviruses* publication either by clicking on the links in the first line, or scrolling down the page to where the Tables of Contents begin.

## The Human Papillomaviruses Compendium On Line

You may now view and print out all or parts of the  $\underline{Human\ Papillomaviruses\ 1994}$  Compendium or the  $\underline{Human\ Papillomaviruses\ 1995}$  Compendium that interest you from this Web site. These two documents consist of  $\sqrt[4]{h_0}$ 

Adobe Acrobat PDF files. "PDF" (Portable Document Format) files retain all the typographical qualities of the original document -- fonts, tables, illustrations. They can be viewed and printed by anyone using a freely available copy of the **Adobe Acrobat** viewer which is available for Macintosh, UNIX, and DOS computers. Your free copy of the Acrobat viewer is available from Adobe Systems.

If you wish to learn more about our use of Acrobat and PDF files click here.

### Hard Copy

The Compendium has been published in paper form and is available to qualified researchers who cannot make use of the electronic hard copy provided here. If you really need a complete paper copy write to us at

The graphic below shows a part of the table of contents of the 1995 book. The pointing hand is about to retrieve an article on mRNA transcript maps from Part III. This link, though it looks the same as others, behaves differently. A copy of the mRNA transcript file is transferred to your local computer and the Adobe Acrobat viewer is automatically launched. Acrobat is a "helper application" used to read files that are in a format which is nonstandard for the WWW.

### PART II. Nucleotide and Protein Alignments of E6, E7, E1, E2, E4, E5, L2, L1, and LCR

Note: conventions and symbols used in the alignments are described in the Introduction.

- Introduction (14 K bytes)
- E6 (68 K bytes)
- E7 (54 K bytes)
- E1 (198 K bytes)
- E2 (158 K bytes)

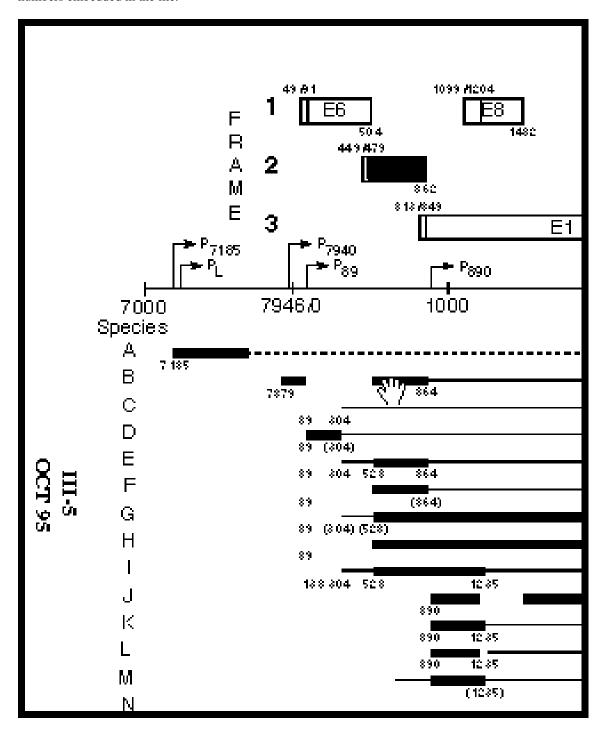
- E4 (72 Kbytes)
   E5 (27 Kbytes)
   L2 (185 Kbytes)
- L1 (171 K bytes)
- L1 Consensus Primer Region (59 K bytes)
- LCR (54 K bytes)

PART III. Genome and RNA transcript maps, reviews of LCR, E6, E7, L1, L2, variation in PVs, patterns and similarities, and recombination.

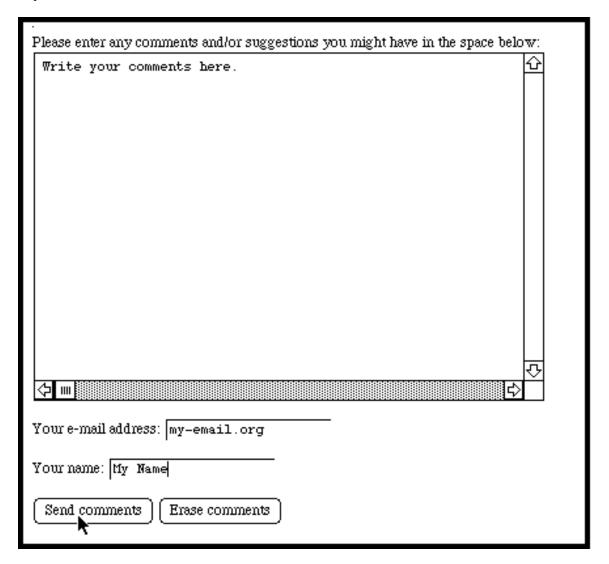
- Genome and mRNA transcript maps (86 K bytes)
   Review of the LCR ( 94 K bytes)
- Overview of PV variants (41 K bytes)
- E6 protein (86 K bytes)
- E7 protein (122 K bytes)
- L1 and L2 proteins (329 K bytes)
- Patterns and similarities (131 K bytes)
- Recombination part 1 (126 K bytes)
- Recombination part 2 (27 K bytes)

PART IV. Nucleotide Sequences of Cellular Proteins Affected by Papillomaviruses

The Acrobat viewer displays the mRNA transcript article (see below), which is written in "portable document format" (PDF), and allows you to print out a perfect copy of it, exactly as it appears in the printed compendium. Acrobat viewer has a search mechanism that allows you to search in a document for words that might lead you to a subject of interest. For example, Part V of the compendium lists hundreds of HPV references. Those references are searchable with Acrobat Viewer. Searches for specific nucleotide sequences in Part I (e.g. AATAAA) will be less useful because of the way GenBank files are formatted, with spaces and sequence numbers embedded in the file.



Our Web site provides users an opportunity to communicate with us from several points on the site. An electronic "comments" form can be activated that allows you to write us a letter to which we promise to respond.



Finally, we have connected our site to several other locations of great utility to biologists as shown below. The link to these sites occurs at the bottom of the home page.

### Links to Hot Web sites

We provide here links to three useful gateways to biological data and tools.

- <u>Entrez</u> browser which allows you to search comprehensive databases at the National Center for Biotechnology Information (NCBI). Entrez allows searches of nucleotide and protein sequence databases and the MEDLINE collection of bibliography information.
- <u>Baylor College of Medicine Home Page</u> provides among other things the Biologist's Control Panel that allows searches of a variety of databases.
- <u>Pedro's Biomolecular Research Tools.</u> A collection of WWW Links to Information and Services Useful to Molecular Biologists.
- National Institue of Allergy and Infectious Diseases (NIAID)

#### Links to Public Health Sources

Our group concentrates on molecular aspects of the human papillomavirus proteins and nucleic acids. We are not qualified to provide information about HPV infections from a medical or public health standpoint. Please do not contact us with health-related questions. For answers to your questions, you might wish to contact the following:

The American Social Health Assocation, whose National Sexually Transmitted Disease Hotline
can be reached at 800/227-8922. They should be able to provide you with information about the
HPV Support Program, which produces a news letter called "HPV News", and coordinates local

**Data Access by FTP** The complete database of sequences and alignments are also available electronically from a server here at Los Alamos called "atlas". In the first section of Part V we described accessing this, and other, information using the WWW. You may also retrieve information from atlas by the "anonymous ftp" method. In the following instructions for accessing atlas, what you actually type at your terminal is shown in **boldface** type [with an explanation shown in square brackets]. What the server responds with is shown indented in italic type.

ftp atlas.lanl.gov [this ftp command connects your computer to atlas]

Name (atlas.lanl.gov): [atlas asks for your user name]

**anonymous** [for offsite users anonymous is the only name that works]

Guest login ok, send ident as password.

Password:

### [enter your e-mail address or your name as ident]

Guest login ok. access restrictions apply.

**cd pub** [change to the directory called pub]

CWD command successful

### cd papilloma

ls [this lists the files and directories inside of papilloma]

SWISS-PROT-files, RoadMap, Alignments, ReadMe, EMBL-files, GenBank-files

pwd [to show your location in the file hierarchy]

"/pub/papilloma" is current directory.

### cd GenBank-files/Human-papilloma [move down 2 directories]

**ls** [all the GenBank files are listed]

**get HPV47.gb** [this copies the file HPV47.gb to your home computer]

**bye** [this disconnects you from the server]

A map of the HPV directories and files on atlas is reproduced on the following page. Many of the directories contain ReadMe files that describe the contents.

### Instructions for Accessing ftp Server "atlas"

In the following instructions, what you actually type at your terminal is shown in **boldface** type [with an explanation shown in square brackets]. What the server responds with is shown indented in *italic* type.

#### ftp atlas.lanl.gov

Name (atlas.lanl.gov):

#### anonymous

Guest login ok, send ident as password.

Password:

#### [enter your e-mail address as ident]

Guest login ok. access restrictions apply.

**cd pub** [change to the directory called pub] *CWD command successful* 

### cd papilloma

ls [this lists the files and directories inside of papilloma]

SWISS-PROT-files, RoadMap, Alignments, ReadMe, EMBL-files, GenBank-files pwd [to show your location in the file hierarchy]

"/pub/papilloma" is current directory.

[all the GenBank files are listed]

**get HPV47.gb** [this copies the file HPV47.gb to your home computer]

bye [this disconnects you from the server]

### **NOVEMBER 1995\* MAP OF HPV DATABASE ON "ATLAS"**

\* Data subject to continuous revision

